(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 054 390 A2

# (12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
**22.11.2000 Bulletin 2000/47**

(51) Int. Cl.[7]: **G10L 15/26**

(21) Application number: **00304117.5**

(22) Date of filing: **16.05.2000**

(84) Designated Contracting States:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE**
Designated Extension States:
**AL LT LV MK RO SI**

(30) Priority: **21.05.1999 US 316643**

(71) Applicant:
**Information Storage Devices, Inc.**
**San Jose, California 95134 (US)**

(72) Inventors:
- **Geilhufe, Michael**
  **Palto Alto, California 94301 (US)**
- **MacMillan, David**
  **Woodside, California 94062 (US)**
- **Barel, Avraham**
  **Doar na shimshon 99782 (IL)**
- **Brown, Amos**
  **Givat smmuhel 99782 (IL)**
- **Bootsma, Karin Lissette**
  **San Jose, California 95134 (US)**
- **Gaddy, Lawrence Kent**
  **San Jose, California 95118 (US)**
- **Pyo, Phillip Paul**
  **San Jose, California 95130 (US)**

(74) Representative:
**Wombwell, Francis et al**
**Potts, Kerr & Co.**
**15, Hamilton Square**
**Birkenhead Merseyside CH41 6BR (GB)**

## (54) Method and apparatus for addressing voice controlled devices

(57) Voice controlled devices with speech recognition have user assignable appliance names and default appliance names to address and control the voice controlled devices. Methods of controling voice controlled devices include addressing a voice controlled device by name and providing a command.

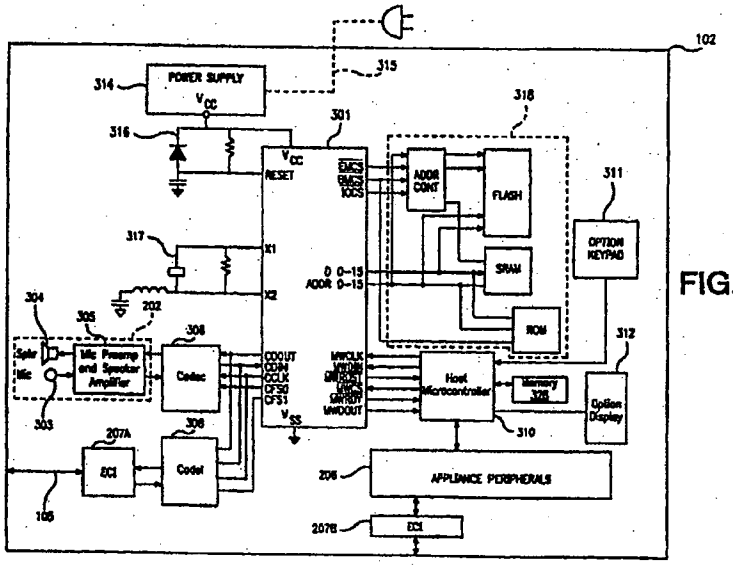



FIG. 3

EP 1 054 390 A2

any user.

**[0007]** In order to achieve high accuracy speech recognition it is important that a voice controlled device avoid responding to speech that isn't directed to it. That is, voice controlled devices should not respond to background conversation, to noises, or to commands to other voice controlled devices. However, filtering out background sounds must not be so effective that it also prevents recognition of speech directed to the voice controlled device. Finding the right mix of rejection of background sounds and recognition of speech directed to a voice controlled device is particularly challenging in speaker-independent systems. In speaker-independent systems, the voice controlled device must be able to respond to a wide range of voices, and therefore can not use a highly restrictive filter for background sounds. In contrast, a speaker-dependant system need only listen for a particular person's voice, and thus can employ a more stringent filter for background sounds. Despite this advantage in speaker dependant systems, filtering out background sounds is still a significant challenge.

**[0008]** In some prior art systems, background conversation has been filtered out by having a user physically press a button in order to activate speech recognition. The disadvantage of this approach is that it requires the user to interact with the voice controlled device physically, rather than strictly by voice or speech. One of the potential advantages of voice controlled devices is that they offer the promise of true hands-free operation. Elimination of the need to press a button to activate speech recognition would go a long way to making this hands-free objective achievable.

**[0009]** Additionally, in locations with a number of people talking, a voice controlled device should disregard all speech unless it is directed to it. For example, if a person says to another person "I'll call John", the cellphone in his pocket should not interpret the "call John" as a command. If there are multiple voice controlled devices in one location, there should be a way to uniquely identify which voice controlled device a user wishes to control. For example, consider a room that may have multiple voice controlled telephones - perhaps a couple of desktop phones, and multiple cellphones - one for each person. If someone were to say "Call 555-1212", each phone may try to place the call unless there was a means for them to disregard certain commands. In the case where a voice controlled device is to be controlled by multiple users, it is desirable for the voice controlled device; to know which user is commanding it. For example, a voice controlled desktop phone in a house may be used by a husband, wife and child. Each would could have their own phonebook of frequently called numbers. When the voice controlled device is told "Call Mother", it needs to know which user is issuing the command so that it can call the right person (i.e. should it call the husbands mother, the wife's mother, or the child's mother at her work number?). Additionally, a voice controlled device with multiple users may need a method to enforce security to protect it from unauthorized use or to protect a user's personalized settings from unintentional or malicious interactions by others (including snooping, changing, deleting, or adding to the settings). Furthermore, in a location where there are multiple voice controlled devices, there should be a way to identify the presence of voice controlled devices. For example, consider a traveler arriving at a new hotel room. Upon entering the hotel room, the traveler would like to know what voice controlled devices may be present and how to control them. It is desirable that the identification process be standardized so that all voice controlled devices may be identified in the same way.

**[0010]** In voice controlled devices, it is desirable to store phrases under voice control. A phrase is defined as a single word, or a group of words treated as a unit. This storing might be to set options or create personalized settings. For example, in a voice-controlled telephone, it is desirable to store people's names and phone numbers under voice control into a personalized phone book. At a later time, this phone book can be used to call people by speaking their name (e.g. "Cellphone call John Smith", or "Cellphone call Mother").

**[0011]** Prior art approaches to storing the phrase ("John Smith") operate by storing the phrase in a compressed, uncompressed, or transformed manner that attempts to preserve the actual sound. Detection of the phrase in a command (i.e. detecting that John is to be called in the example above) then relies on a sound-based comparison between the original stored speech sound and the spoken command. Sometimes the stored waveform is transformed into the frequency domain and / or is time adjusted to facilitate the match, but in any case the fundamental operation being performed is one that compares the actual sounds. The stored sound representation and comparison for detection suffers from a number of disadvantages. If a speaker's voice changes, perhaps due to a cold, stress, fatigue, noisy or distorting connection by telephone, or other factors, the comparison typically is not successful and stored phrases are not recognized. Because the phrase is stored as a sound representation, there is no way to extract a text-based representation of the phrase. Additionally, storing a sound representation results in a speaker dependent system. It is unlikely that another person could speak the same phrase using the same sounds in a command and have it be correctly recognized. It would not be reliable, for example, for a secretary to store phonebook entries and a manager to make calls using those entries. It is desirable to provide a speaker independent storage means. Additionally, if the phrases are stored as sound representations, the stored phrases can not be used in another voice controlled device unless the same waveform processing algorithms are used by both voice controlled devices. It is desirable to recognize spoken phrases and store them in a representation such that, once stored, the phrases can be used for speaker independent recognition and can be used by multiple voice controlled devices.

**[0012]** Presently computers and other devices communicate commands and data to other computers or devices using modern, infrared or wireless radio frequency transmission. The transmitted command and/or data are usually of

FIGs. 10A-10C are flow charts of the "GETRESPONSE" function for the standard voice user interface of the present invention.

FIG. 11 is a flow chart of the "GETRESPONSEPLUS" function for the standard voice user interface of the present invention.

FIG. 12 is a flow chart of the "LISTANDSELECT" function for the standard voice user interface of the present invention.

FIG. 13 is a block diagram of a pair of voice controlled devices communicating using the standard voice user interface of the present invention.

Like reference numbers and designations in the drawings indicate like elements providing similar functionality.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

**[0018]** In the following detailed description of the present invention, numerous specific details are set forth in order to provide a thorough understanding of the present invention. However, it will be obvious to one skilled in the art that the present invention may be practiced without these specific details. In other instances well known methods, procedures, components, and circuits have not been described in detail so as not to unnecessarily obscure aspects of the present invention.

**[0019]** The present invention includes a method, apparatus and system for standard voice user interface and voice controlled devices. Briefly, a standard voice user interface is provided to control various devices by using standard speech commands. The standard VUI provides a set of core VUI commands and syntax for the interface between a user and the voice controlled device. The core VUI commands include an identification phrase to determine if voice controlled devices are available in an environment. Other core VUI commands provide for determining the names of the voice controlled devices and altering them.

**[0020]** Voice controlled devices are disclosed. A voice controlled device is defined herein as any device that is controlled by speech, which is either audible or non-audible. Audible and non-audible are defined herein later. A voice controlled device may also be referred to herein as an appliance, a machine, a voice controlled appliance, a voice controlled electronic device, a name activated electronic device, a speech controlled device, a voice activated electronic appliance, a voice activated appliance, a voice controlled electronic device, or a self-identifying voice controlled electronic device.

**[0021]** The present invention is controlled by and communicates using audible and non-audible speech. Speech as defined herein for the present invention encompasses a) a signal or information, such that if the signal or information were passed through a suitable device to convert it to variations in air pressure, the signal or information could be heard by a human being and would be considered language, and b) a signal or information comprising actual variations in air pressure, such that if a human being were to hear the signal, the human would consider it language. Audible speech refers to speech that a human can hear unassisted. Non-audible speech refers to any encodings or representations of speech that are not included under the definition of audible speech, including that which may be communicated outside the hearing range of humans and transmission media other than air. The definition of speech includes speech that is emitted from a human and emitted from a machine (including machine speech synthesis, playback of previously recorded human speech such as prompts, or other forms).

**[0022]** Prompts which are communicated by a voice controlled device and phrases which are communicated by a user may be in languages or dialects other than English or a combination of multiple languages. A phrase is defined herein as a single word, or a group of words treated as a unit. A user, as defined herein, is a human or a device, including a voice activated device. Hence "a user's spoken phrase", "a user issuing a command", and all other actions by a user include actions by a device and by a human.

**[0023]** Voice controlled devices include some type of speech recognition in order to be controlled by speech. Speech recognition and voice recognition are used synonomously herein and have the same meaning. Preferably, speeker independent speech recognition systems are used to provide the speech recognition capablity of the voice controlled devices. Speaker independent speech recognitions systems are responsive to speaker-independent representations of speech. In the prefered embodiment, a speaker-independent representation of speech is a phonetic representation of speech. However, other speaker-independent representations of speech may also be used in accordance with the present invention.

**[0024]** In order to gain access to the full functionality of a voice controlled device with the present invention, a user must communicate to the voice controlled device one of its associated appliance names. The appliance name may include one or more default names or one or more user-assignable names. A voice controlled device may have a plurality of user-assignable names associated with it in order to provide personalized functionality to each user.

lap in time], they both must back off for a new randomly selected silence delay, but this time the delay must be of up to twice the length of the previous silence delay, but not to exceed 16 seconds.

**[0029]** In order to restrict which voice controlled devices respond to an identification phrase, a user may include a voice controlled device's name in the identification phrase. For example, one could say "Socrates are you out there?" to see if a voice controlled device named Socrates was nearby. Similarly, one could say "Clock are you out there" which would cause all voice controlled devices with an appliance name of Clock (whether a default appliance name or a user appliance name) to respond. A possible variation is that voice controlled devices may respond with some response other than their names, as for example, might be needed for security reasons.

**[0030]** A voice controlled device may use both visual and acoustic identification methods. For example, even though a speech recognition engine is continuously on, it may still display the visual logo and / or other visual identifier. Similarly, in a voice controlled device that requires manual activation of the speech engine, once enabled, the engine could then be responsive to the command "What is out there?"

**[0031]** In another aspect of the present invention, the initial storage of a user's spoken phrase (for example, when making a new phonebook entry under voice control) is processed by the speaker-independent speech recognition engine of the voice controlled devices. This engine returns a speaker-independent phonetic representation of the phrase. This speaker-independent phonetic representation is what is stored.

**[0032]** When a command is issued by a user, it is also processed by the speaker-independent speech recognition engine of the present invention. This could be the same speaker-independent engine use for storing the original entries, or a completely different speaker-independent engine. In either case, the engine returns a speaker-independent phonetic representation of the command sequence. This speaker-independent phonetic representation can be compared to earlier stored phonetic representations to determine whether the command is recognizable.

**[0033]** By converting both the stored spoken entries and any commands to speaker-independent phonetic representation a number of advantages are provided.

- Recognition will be reliable even if the user's voice has changed, perhaps due to a sickness, stress, fatigue, transmission over a noisy or distorting phone link, or other factors that might change a human user's or machine user's speech. Text-based information can be stored and then recognized.
- Recognition will be reliable even if some other user had stored the original voice phrase.
- Recognition can be speaker-independent, even for user-stored commands and phrases.
- Stored entries originating from text sources and from different speakers can all be combined and reliably for recognition.
- The use of speaker-independent phonetic representations facilitates upgrading to improved recognition engines as they become available. Improved speech recognition engines can use existing stored information without impacting reliability or requiring re-storage, since all stored entries are held in phonetic form. New information stored using the improved speech recognition engines can be used on equipment with older recognition engines. Old and new generations of equipment can interoperate without prior coordination by using phonetic representations. This allows, for example, two PDAs to exchange voice-stored phonebook entries and provide reliable recognition to the new users of that information. Finally, there are no legacy restrictions to hold back or restrict future development of speaker-independent recognition engines as long as they can create phonetic representations, unlike waveform-storage based systems, which must always be able to perform exactly the same legacy waveform transformations.

VOICE CONTROLLED DEVICES

**[0034]** Referring now to FIG. 1A, environment 100 is illustrated. Environment 100 may be any communication environment such as an office, a conference room, a hotel room, or any location where voice controlled devices may be located. Within environment 100, there are a number of human users 101A-101H, represented by circles. Also within the environment 100, are voice controlled devices 102A-102H, represented by squares and rectangles, each operationally controlled by the standard voice user interface (VUI) of the present invention. Voice controlled devices 102A-102E, represented by rectangles, are fixed within the environment 100. Voice controlled devices 102F-102H, represented by squares, are mobile voice controlled devices that are associated with human users 101F-101H respectively. Voice controlled devices 102A-102H may be existing or future devices. Voice controlled devices 102A-102E may be commonly associated with a user's automobile, home, office, factory, hotel or other locations where human users may be found. Alternatively, if the voice controlled devices 102A-102E are to be controlled by non-audible speech, voice controlled devices may be located anywhere.

**[0035]** In the present invention, the standard VUI allows a user to associate a user-assignable name with these voice controlled devices 102A-102H. The user-assignable name of the voice controlled device may be generic such as telephone, clock, or light. Alternatively, the name may be personalized such as those ordinarily given to humans such as John, Jim, or George. In either case, the voice controlled devices 102A-102H while constantly listening will not

to SRS 204 and SRS 204 is coupled to APCC 206 . In the voice controlled device 102I, ACI 202 is its primary means of speech communication.

**[0039]** Voice controlled device 102J includes ACI 202, SRS 204, APCC 206, communications interface (ECI) 207, and connection 208. ACI 202 is coupled to SRS 204. APCC 206 is coupled to SRS 204. ECI 207 couples to SRS 204 and connection 208 couples to the ECI 207. Voice controlled device 102J can alternatively communicate using speech or voice communication signals through ACI 202 or ECI 207. Voice controlled device 102K includes ACI 202, SRS 204, APCC 206, and an antenna 209.

**[0040]** Voice controlled device 102K can communicate using audible speech signals through the ACI 202 or using encoded speech signals through the ECI 207. ECI 207 couples to APCC 206. ECI 207 also couples to Connection 212. Connection 212 could, for example, be an antenna or infrared port. Voice controlled device 102L also includes an ACI 202, SRS 204, APCC 206, and an antenna 209. ACI 202 couples to SRS 204. SRS 204 couples to APCC 206. Antenna 209 couples to APCC 206. Voice controlled device 102L can communicate by means of ACI 202 and APCC 206 through antenna 209.

**[0041]** Voice controlled device 102M includes an APCC 206, SRS 204, an ECI 207, and connection 210. Connection 210 may be a wired or wireless connection, including an antenna. SRS 204 couples to APCC 206 and also to ECI 207. Connection 210 couples to ECI 207. Voice controlled device 102M can communicate via ECI 207 over connection 210.

**[0042]** The APCC 206 represents the elements of the voice controlled device 102 that are to be controlled. For example, in the case of white goods, the items to be controlled may be temperature, a time setting, a power setting, or a cycle depending on the application. In the case of consumer electronics, the APCC 206 may consist of those items normally associated with buttons, switches, or knobs. In the case of telephone products, the APCC 206 may represent the buttons, the dials, the display devices, and the circuitry or radio equipment for making wired or wireless calls. In the case of automobile systems, the APCC 206 may represent instrumentation panels, temperature knobs, navigational systems, the automobile radios channels, volume, and frequency characteristics.

**[0043]** Referring now to FIG. 3, the voice controlled device 102 is illustrated. Voice controlled device 102, illustrated in FIG. 3, is exemplary of the functional blocks within voice controlled devices described herein. Voice controlled device 102 includes the ACI 202, the APCC 206 and the SRS 204. The voice controlled device 102 may also have an ECI 207 such as ECI 207A or ECI 207B.

**[0044]** The ACI 202 illustrated in FIG. 3 includes microphone 303, speaker 304, and amplifiers 305. The SRS 204 as illustrated in FIG. 3 includes the voice communication chip 301, coder/decoder (CODEC) 306 and 308, host microcontroller 310, power supply 314, power on reset circuit 316, quartz crystal oscillator circuit 317, memory 318, and memory 328. The SRS 204 may optionally include an AC power supply connection 315, an optional keypad 311 or an optional display 312. For bidirectional communication of audible speech, such as for local commands, prompts and data, the speech communication path is through the VCC 301, CODEC 306, and the ACI 202. For bidirectional communication of non-audible speech, such as for remote commands, prompts and data, the non-audible speech communication path is through the VCC 301, CODEC 308, ECI 207A or the VCC 301, host microcontroller 310, APCC 206, and ECI 207B. The ECI 207 may provide for a wired or wireless link such as through a telephone network, computer network, internet, radio frequency link, or infrared link.

**[0045]** Voice communication chip 301 provides the voice controlled device 102 with a capability of communication via speech using the standard voice user interface of the present invention. Microphone 303 provides the voice controlled device 102 with the capability of listening for audible speech, such as voice commands and the device's appliance names. Microphone 303 may be a near field or far field microphone depending upon the application. For example, near field microphones may be preferable in portable cell phones where a user's mouth is close while far field microphones may be preferable in car cell phones where a user's mouth is a distance away. Speaker 303 allows the voice controlled device 102 to respond using speech such as for acknowledging receipt of its name or commands. Amplifiers 305 provides amplification for the voice or speech signals received by the microphone 303. Additionally, the amplifiers 305 allow amplification of representations of voice signals from the CODEC 306 out through the speakers 303 such that a human user 101 can properly interface to the voice controlled device 102.

**[0046]** Microphone 303 and Speaker 304 are each transducers for converting between audible speech and representations of speech. CODEC 306 encodes representations of speech from the ACI 202 into an encoded speech signal for VCC 301. In addition, CODEC 306 decodes an encoded speech signal from the VCC 301 into an representation of speech for audible communication through the ACI 202.

**[0047]** Alternatively, non-audible speech signals may be bi-directionally communicated by the voice controlled device 102. In this case, VCC 301 provides encoded speech signals to CODEC 308 for decoding. CODEC 308 decodes the encoder speech signal and provides it to the ECI 207A for communication over the connection 105. Speech signals may be received over the connection 105 and provided to the ECI 207A. The ECI 207A couples the speech signals into the CODEC 308 for encoding. CODEC 308 encodes the speech signals into encoded speech signals, which are coupled into the VCC 301.

cycle. The MICROWIRE Interface 428 allows serial communication with the host microcontroller 310. The Master MICROWIRE controller 430 allows interface to serial flash memory and other peripherals. The reset and configuration block 432 controls definition of the environment of the voice communication chip 301 during reset and handles software controlled configurations. Some of the functions within the voice communication chip 301 are mutually exclusive. Selection among the alternatives is made upon reset or via a Module Configuration register. The clock generator 434 interfaces to the quartz crystal oscillator circuit 317 to provide clocks for the various blocks of the voice communication chip including a real-time timer. The clock generator can also be used to reduce power consumption by setting the voice communication chip 301 into a powerdown mode and returning it into normal operation mode when necessary. When the voice communication chip 301 is in power-down mode, some of its functions are disabled and contents of some registers are altered. The watchdog timer 436 generates a non-maskable interrupt whenever software loses control of the processing units 402 and at the expiration of a time period when the voice communication chip 301 is in a power-down mode.

STANDARD VOICE USER INTERFACE

**[0058]** Similar to computer operating systems providing a GUI, the standard voice user interface (VUI) can be thought as being provided by a standard VUI operating system code. The standard VUI operating across a wide array of voice controlled devices allows a user to interface any one of the voice controlled devices including those a user has never previously interacted with. Once a user is familiar with the standard VUI, they can walk up to and immediately start using any voice controlled device operating with the standard VUI. The standard VUI operating system code has specific standardized commands and procedures in which to operate a voice controlled device. These standardized commands and procedures are universal to machines executing the standard VUI operating system code. Voice controlled application software, operating with the standard VUI operating system code, can be written to customize voice controlled devices to specific applications. The voice controlled application software has voice commands specific to the application to which the voice controlled device is used. A particular voice controlled device may also have additional special features that extend the core capabilities of the standard VUI.

**[0059]** Some of the standard VUI functionality in the core VUI include a way to discover the presence of voice controlled devices, a core common set of commands for all voice controlled devices, a way to learn what commands (both core commands and appliance-specific commands) the voice controlled device will respond to, a vocalized help system to assist a user without the use of a manual or display, a way to personalize the voice controlled device to a user with user assignable settings, security mechanisms to control use of voice controlled devices to authorized users and protect user assignable settings and information from other users, and standard ways for a user to interact with voice controlled devices for common operations (e.g. selecting yes or no, listing and selecting items from a list of options, handling errors gracefully, etc.). The standard VUI includes an API (Applications Programming Interface) to allow software developers to write custom voice controlled applications that interface and operate with the standard VUI and extend the voice controlled command set.

**[0060]** Referring now to FIG. 5, a block diagram illustrates the Software 500 for controlling Voice Controlled Device 102 and which provides the standard VUI and other functionality. The Software 500 includes Application Code 510, a VUI software module 512 and a Vocabulary 524. Application code 510 may be further modified to support more than one application, representing multiple application code modules, to provide for further customization of a voice controlled device 102. The Vocabulary 524 contains the phrases to be detected. The phrases within the Vocabulary are divided into groups called Topics, of which there may be one or more. In Figure 5, the Vocabulary 524 consists of two Topics, Topic 551 and Topic 552.

**[0061]** Typically, Application Code 510 interfaces to the VUI software 512 through the Application Programming Interface (API) 507. The VUI software 512 provides special services to the Application Code 510 related to voice interface, including recognition and prompting. The interrelationship between the VUI software 512 and the application code 510 is analogous to that between Microsoft's MS Windows and Microsoft Word. Microsoft Windows provides special services to Microsoft Word related to displaying items on a screen and receiving mouse and keyboard inputs.

**[0062]** Generally, the Application Code 510 may be stored in host memory and executed by the host microcontroller 310. However, the functionality of the host microcontroller 310 can be embedded into the VCC 301 such that only one device or processor and one memory or storage device is needed to execute the code associated with the software 500.

**[0063]** All phrases that can be recognized, including those phrases for the core and application specific commands, are included in the Vocabulary 524. The VUI software module 512 can directly access the vocabulary phrases, for example for use during recognition. The VUI software module 512 can also processes tokens. Tokens abstractly relate to the phrases within the Topics 551-552. Tokens are integer numbers. For example, the phrase for 'dial' might have a token value of '5', and the phrase for 'hang-up' might have a token value of '6'. There is a token value assigned to every phrase that can be recognized. Because the VUI software module 512 can process tokens related to the vocabulary file 524, it can refer to phrases without having to directly access them. This makes it possible to change languages (from

ground speech may still be present. The ⟨name⟩ is the appliance name associated with a voice controlled device 102. The ⟨command⟩ is an operation that a user wants performed. The ⟨modifiers & variables⟩ consist of additional information needed by some commands. The SRS 204 recognizes the elements in their syntax in order for a user to control voice controlled devices.

**[0072]** Most voice controlled devices will continuously listen for the voice command sequence. When a voice controlled device hears its ⟨name⟩, it knows that the following ⟨command⟩ is intended for it. Since each user has a different ⟨name⟩ for a voice controlled device, the ⟨name⟩ also uniquely identifies the user, allowing the voice controlled device to select that user's personalization settings. Commands include core VUI commands included with all voice controlled devices, and commands specific to a given application, all of which are stored within the vocabulary 524.

**[0073]** Requiring ⟨silence⟩ before detection of ⟨name⟩ helps prevent false detection of ⟨name⟩ during normal conversational speech (i.e. during periods when the user is speaking conversationally to other users and not to the voice controlled device). In all cases, the duration of ⟨silence⟩ can be configured by the manufacturer and can range from 0 (no ⟨silence⟩ required) to a second or more. Typically it will be about a quarter of a second.

**[0074]** Examples of voice command sequences that might be used with a voice controlled device such as a telephone named Aardvark include "Aardvark Call The Office", "Aardvark Dial 1-800-55-1212", and "Aardvark Hang-up". (In the command examples and descriptions provided, for the sake of brevity the ⟨silence⟩ is often not shown, and even where it is shown or described, the option always exists of a manufacturer choosing to use a silence duration of zero.)

**[0075]** There are two special cases where the command syntax is permitted to differ from the general syntax. The first special case is in voice controlled devices that do not continuously listen for ⟨silence⟩⟨name⟩. For example, in some battery operated applications, power consumption limitations may require the VCC 301 in the voice controlled device 102 to be powered down during idle periods. Another example is a voice controlled device located where false recognition of a name would have undesirable results, for example, a desktop phone in a conference room during a presentation. A third example is voice controlled devices where there is a high risk of false recognition, for example, where multiple conversations can be heard.

**[0076]** For these types of situations, an alternate command syntax is used in conjunction with a button or switch of some type. The first alternate command syntax is:

⟨activation of a switch⟩ ⟨silence (optional)⟩ ⟨name⟩ ⟨command⟩ ⟨modifiers & variables⟩.

In this syntax, the ⟨activation of a switch⟩ means the user presses a button or performs some other mechanical act (e.g. opening a flip-style cell phone) to activate the recognition capability.

**[0077]** A second special case is where the user normally enters a series of commands in quick succession. For these cases, the user can identify themselves once to the voice controlled device using a password protection method, or by issuing a command that includes the voice controlled device's appliances ⟨name⟩, and thereafter continue entering commands. The second alternate command syntax (in this example, for three successive commands) is:

⟨silence⟩ ⟨name⟩ ⟨command⟩ ⟨modifiers & variables as needed⟩
⟨silence⟩ ⟨name (optional)⟩ ⟨command⟩ ⟨modifiers & variables as needed⟩
⟨silence⟩ ⟨name (optional)⟩ ⟨command⟩ ⟨modifiers & variables as needed⟩

With this syntax, the user can issue a series of commands without having to constantly repeat the voice controlled device's appliances ⟨name⟩. However, the user is permitted to say the ⟨name⟩ at the start of a command. Note that in this syntax, the ⟨silence⟩ is required to properly recognize the spoken ⟨name⟩ or ⟨command⟩.

**[0078]** When either of the first or second alternate syntaxes is used, it is desirable to ensure that if a new user starts working with the voice controlled device, they are properly identified. This can be ensured by explicitly requiring the ⟨name⟩ after a period of inactivity or after power-up of the voice controlled device or other similar protocol.

STANDARD CORE VUI COMMANDS

**[0079]** There are a number of standard core commands included in the vocabulary 524 of voice controlled devices 102 operating using the standard VUI. FIGs. 6A-8 illustrate the syntax of the following commands.

**[0080]** Referring to FIG. 6A, at start 600, the appliance name, ⟨name⟩, of a voice controlled device is usually spoken prior to a command. Any of the voice controlled device's appliances names can be spoken whenever the voice controlled device is listening for a command. If the ⟨name⟩ is not followed by a command within some period of time, the voice controlled device will go back to return to start 600 in its original idle state. This is indicated by the solid box Silence of N seconds. N in this case is a programmable value usually application dependent and assigned by the voice controlled device manufacturer. After supplying the appliance name, a user is granted access to further commands of the standard VUI operating on the voice controlled device at 601.

4 to (for example) Doggone. If the user attempted to change a fifth user-assignable name in sequence with the command ("Telephone change your name "), it would result in an error message because all available user-assignable appliance names were assigned. Note that the voice controlled device always responds to the factory programmed name, even if all user-assigned names are defined. Accordingly, in this example of a fifth attempt, the voice controlled device still recognizes the "Telephone" factory programmed name - it is just unable to assign a fifth new user-assignable appliance name.

**[0086]** An existing user-assignable appliance name can also be changed with the "Change Your Name " command. Continuing the above example, "Aardvark change your name " would alter the appliance's name for the first user (for example, it could be changed to Platypus), and leave the other three user names unchanged. Similarly, "Platypus change your name " followed by a dialog to set the name to "Telephone" would reset the first user name to the factory-programmed default.

Identification of Voice Controlled Devices

**[0087]** As voice controlled devices proliferate, it is important that users be capable of readily identifying what, if any, voice controlled devices are present when they enter a new environment. For example, a user walks into a hotel room that has a number of devices. In order to use them a user needs to know which devices are voice controlled devices. Additionally a user needs to know the appliance names in order to properly control them. Beside being audibly identified, voice controlled devices can be identified visually as well as by using a logo signifying a voice controlled device utilizing the standard VUI.

**[0088]** Acoustic identification works when voice controlled devices are actively listening for recognizable commands. In most cases, this means the voice controlled device is constantly listening and attempting recognition. Typically, these voice controlled devices will be AC powered, since the power drain from continuous recognition will be unacceptable for most battery operated voice controlled devices. Referring to FIG. 6A and 6C, the acoustic identification is accomplished by a user communicating an identification phrase to command the voice controlled device. The identification phrase "What Is Out There?" or some other suitable identification phrase may be used for causing the voice controlled devices to identify themselves.

**[0089]** The syntax of the standard VUI identification phrase is:

(silence) What Is Out There?

In response to this query, any voice controlled device that hears the question must respond. The typical voice controlled devices response is a random delay of up to 2 seconds of relative silence, followed by a beep (the standard signal) , and the response "You can call me (name)", where (name) is the factory-programmed name that can be used to address the voice controlled device. In the telephony voice controlled device example described above, a response might be "(beep) You can call me Telephone."

**[0090]** Referring to FIG. 6C, during the random delay of up to 2 seconds, each responding voice controlled device listens for another voice controlled device's response (specifically, for another voice controlled device's beep). In the event another voice controlled device starts responding (as evidenced by a beep) during this silence period, the listening voice controlled device must restart its silence timing after the responding voice controlled device finishes. In the event two voice controlled devices start responding at the same time (overlapping beeps), they both must back off for a new randomly selected silence delay. However, this time the random delay may be greater than the first, up to twice the length of the previous silence delay. In any event, the delay should not exceed 16 seconds. Additional back off periods for further conflict resolution is provided if other voice controlled devices respond.

**[0091]** Referring to FIG. 6A, the syntax of the Request User-Assignable Names command is:

(name) Tell Me Your Name
or
(name) Tell Me Your Names

If security permits, any user-programmed (name) or the default (name) can be used. The Request User-Assignable Names command is used to ask a voice controlled device to list all the user-programmed (names) that it will respond to. If security permits, the voice controlled device communicates each use-programmed name in a list fashion. Between each user-assigned name it pauses for a moment. During this pause a user may communicate a command to the voice controlled device and it will be executed as if given with that user-programmed (name). For example consider the telephony voice controlled device example above. The command "Telephone Tell Me Your Name" provided after a pause will cause the telephone to respond by saying "I have been named Aardvark, (pause) Barracuda (pause), Coyote (pause), and Doggone (pause)." During the pause that followed the voice controlled device saying "Coyote", a user may say "Call

**[0102]** The syntax of the Call command is:

⟨name⟩ Call ⟨voicetag⟩
or
⟨name⟩ Call ⟨digits⟩

The Call command is used to dial a specific phone number, expressed either as a series of digits or as a phonebook voicetag. The ⟨digits⟩ can be any list of numeric digits. The telephony voice controlled device allows for the synonyms "oh" for zero, and "hundred" for zero-zero to be enabled. The sequence of ⟨digits⟩ can contain embedded pauses. However, if a pause exceeds a programmable duration, the sequence is terminated and the command executed after recognition of a pause that exceeds a duration set by the system designer. The telephony voice controlled device response to a Call command should be "Calling ⟨digits⟩" or "Calling ⟨voicetag⟩" with the recognized digits or recognized voicetag voiced to verify accurate recognition. The "Cancel" command can be used to cancel the calling operation in the event of misrecognition.

**[0103]** The syntax of the Dial command is:

⟨name⟩ Dial ⟨voicetag⟩
or
⟨name⟩ Dial ⟨digits⟩

The Dial command is the same as the Call command.

**[0104]** The syntax of the Answer command is:

⟨name⟩ Answer

This command is used to answer an incoming call. The response prompt is "Go ahead".

**[0105]** The syntax of the Hangup command is:

⟨name⟩ Hangup

This command is used to hangup an active call. The response prompt is a high-pitched beep.

**[0106]** The syntax of the Redial command is:

⟨name⟩ Redial

This command is used to redial a number. The response is "Redialing ⟨digits⟩" or "Redialing ⟨voicetag⟩", depending on whether the previous Call or Dial command was to ⟨digits⟩ or a ⟨voicetag⟩. If there was no earlier call made, the response is "Nothing to redial".

**[0107]** The syntax of the Store command is:

⟨name⟩ Store

The Store command is in the phonebook submenu and is used to add a new voicetag.

**[0108]** The syntax of the Delete command is:

⟨name⟩ Delete

The Delete command is in the phonebook submenu and is used to delete a voicetag.

**[0109]** The syntax of the Mute command is:

⟨name⟩ Mute

This command mutes the microphone. The response by the voice controlled device is "Muted".

**[0110]** The syntax of the Online command is:

⟨name⟩ Online

This command unmutes the microphone. The response is "Online".

| | | |
|---|---|---|
| name" | number" | |
| "Please repeat the new name" | "The number for <voicetag> is <digits>. Is this correct?" | "nine" |
| "Please say the number for <voicetag>" | "The number for <voicetag> has been stored" | "zero" |
| "That name is not in the phone book" | "Do you want to store it now?" | "hundred" |
| | "Muted" | "Nothing to redial" |
| | | "Star" |
| | | "Flash" |
| | | "Pound" |

**[0112]** In addition to these prompts, the voice controlled devices can generate a number of different tones or beeps. These include a medium pitch beep (e.g. 200 millisecond, 500 Hz. sine wave), a low pitched beep (e.g. a buzzer sound or 250 millisecond, low frequency beep signifying erroneous entry) and a high pitched beep (e.g. 200 milliseconds, 1200 Hz. sine wave). Other sounds are possible and would be within the intended scope of the present invention.

Vocabulary For Telephone Answering Voice Controlled Device

**[0113]** In addition to the forgoing, application specific commands for the standard VUI enable a user to interface to a telephone answering voice controlled device using voice commands. A user can manage message functions and obtain remote access from a telephone answering voice controlled device without using a keypad. The following lists the additional voice commands to be included in the vocabulary 224 for telephone answering voice controlled device.

| | | |
|---|---|---|
| <name> Play new | <name> Rewind <n> | <name> Stop |
| <name> Play all | <name> Record Greeting | <name> Play Greeting |
| <name> Delete this | <name> Record message | <name> Room monitor |
| <name> Delete all messages | <name> Answer On | <name> Password <password phrase> |
| <name> Forward <n> | <name> Answer Off | |

103. If the recognition of the response was not successful, the array is two elements long. The first element is set to zero and the second element indicates the type of error that occurred. In this case, Element 1 is set to 0 indicating that an error was detected. Element 2 is set to 17 indicating that a response was not detected in the allowed time (Timeout error) or 18 indicating that a response was detected, but it was not recognizable (out-of-vocabulary-word error). The array returned for a timeout error is two elements long with values 0, 17 and the array returned for an out-of-vocabulary-word error is two elements long with values 0, 18.

**[0119]** Referring to FIG. 11, GETRESPONSEPLUS user interface function plays a Prompt to a user that solicits a response and waits for the response. GETRESPONSEPLUS is similar to GETRESPONSE in that it plays a Prompt for the user and then waits for a spoken response. However, GETRESPONSEPLUS includes the capability to play prompts to recover from error situations where the user has not spoken or has excessive noise in the background. GETRESPONSEPLUS listens for a spoken response that matches the topics in TopicList. GETRESPONSEPLUS either returns an array of recognized tokens, or an error indicator. The parameters for GETRESPONSEPLUS are Initial_Prompt, Timeout, STS_Sound, TopicList, MaxTries, Intervene_Prompt, Repeat_Prompt, and the Help_Prompt. The Initial_Prompt parameter is the initial prompt to be played to a user to solicit a response. The TimeOut parameter is the number of milliseconds to wait for a response before flagging that a response was not detected. The STS_Sound prompt is a sound or prompt to be played if user speaks before Prompt finishes playing. Typically, STS_Sound prompt will be a short tone or beep sound rather than a spoken phrase. The parameter TopicList is the vocabulary subset for the list of topics which the SRS 204 should use to identify the spoken response. The MaxTries parameter is the maximum number of times GETRESPONSEPLUS will re-prompt the user in an effort to get a good recognition. If recognition does not occur after MaxTries, GETRESPONSEPLUS will return and indicate an error. The Intervene_Prompt parameter is a prompt played to ask the user to repeat himself (e.g. "There was too much noise. Please repeat what you said."). This prompt is played when there was too much noise during the previous recognition attempt. The Repeat_Prompt parameter is the prompt played to ask the user to repeat what was just said (e.g. "Please repeat what you said"). This prompt is used when a spoke-too-soon error occurred. The Help_Prompt parameter is the prompt played when the user seems to need further instructions, including when the user says nothing. The voice controlled device returns a pointer to an integer array upon completion of the user interface function. If the recognition of a response associated with the TopicList was successful, the first element in the array is the number of tokens returned and the following elements in the array are the tokens for each identified speech element (one or more words). Element 1 is n the Number of tokens returned. Elements 2 through n+1 are the Token values for each speech element recognized. For example, consider the phrase "Telephone Dial Office". If the token value for the speech element "Telephone" is 7, for the speech element "Dial" is 12, and for the speech element "Office" is 103, then if they are all recognized successfully, the complete array returned would be four elements long with the values 3, 7, 12, 103. If recognition was not successful, the array is four elements long. The first element is zero. The second element indicates the most recent type of error that occurred. The third through fifth elements indicate the number of times each type of error occurred between when GETRESPONSEPLUS was called to when GETRESPONSEPLUS returned. In this case Element 1 has a value of 0 indicating that an error was detected. Element 2 has a value of 17 indicating that a response was not detected in the allowed time (TimeOut error) or 18 indicating that a response was detected, but it was not recognizable (out-of-vocabulary-word error) or 19 indicating that a spoke-to-soon error was detected. Element 3 has a value of x indicating the number of times a TimeOut error was detected. Element 4 has a value of y indicating the number of times an out-of-vocabulary-word error was detected. Element 5 has a value of z indicating the number of times a spoke-too-soon error was detected.

**[0120]** Referring to FIG. 12, LISTANDSELECT user interface function first plays a Prompt. Then it plays each prompt in array ListOfMenuPrompts, pausing after each for a PauseTime. During these pauses, the recognizer listens for a spoken response that matches the topics in TopicList. LISTANDSELECT either returns an array of recognized tokens, or an error indicator. The parameters for LISTANDSELECT include Initial_Prompt, Timeout, STS_Sound, TopicList, ListOfMenuPrompts, PauseTime, and the Help_Prompt. The Initial_Prompt parameter is the initial prompt to be played to the user. The TimeOut parameter is the number of milliseconds to wait for a response, after playing all the prompts in ListOfMenuPrompts, or before flagging that a response was not detected. The STS_Sound parameter is the sound or prompt to be played if user speaks before a prompt finishes playing. Typically, STS_Sound will be a short tone or beep sound rather than a spoken phrase. The parameter TopicList is the vocabulary subset for the list of topics which the SRS 204 should use to identify the spoken response. The ListOfMenuPrompts parameter is an array of prompts which will be played one at a time. The first element in the array is a count of the number of prompts in ListOfMenuPrompts. The PauseTime parameter is the time to pause after playing each prompt in ListOfMenuPrompts. The PauseTime parameter has a value in milliseconds. The Help_Prompt parameter is the prompt played when the user seems to need further instructions, including when the user says nothing. The voice controlled device returns a pointer to an integer array upon completion of the user interface function. If recognition was successful, the first element in the array is the number of tokens returned, and the following elements in the array are the tokens for each identified speech element (one or more words). Element 1 has a value of n indicating the number of tokens returned. Elements 2 through

Machine Requests to Humans

**[0125]** Machines can ask humans to do things. Any request should be polite. For example, a voice activated cellular telephone might ask to be placed in its charger when its batteries are running low. Humans should always have the option to refuse a machine's request, and the machine should politely accept that, unless the machine considers the situation threatening to human life or valuable data, in which case its protests can be more urgent.

Machines That Use the Telephone On Their Own

**[0126]** If a voice controlled device answers the telephone, or places a call to a human user, it should clearly identify itself as a machine if there is any risk of it being considered human.

Recording User Speech

**[0127]** No machine should record or transcribe a human user's conversations unless those humans present are aware that this is occurring.

Volume Levels

**[0128]** Machines should modulate their volume levels in response to ambient noise levels, unless specifically overridden by a human. Machines should be sensitive to when humans want them to be silent (for example, when humans are sleeping). Machines shouldn't babble needlessly, and should permit a user barge-in as a means to silence them.

Machine- to-Machine Communication

**[0129]** FIG. 13 is a block diagram of a pair of voice controlled devices 102M and 102N (each also referred to as a machine) communicating, neither, one or both of which could be using the standard voice user interface 500 of the present invention in the communication environment 1300. Voice controlled devices can talk to each other to find out what other voice controlled devices are present, what kinds of information they understand, and to exchange information. For example, a voice controlled TV may ask a voice controlled VCR about necessary settings for it to operate. Machine-to-machine communication between voice controlled devices occurs in both audible and non-audible formats. Essentially, machine-to-machine communication using speech may occur over any speech-compatible media, including sound waves through air, conventional telephone links, Internet voice links, radio voice channels, and the like. Machine-to-machine communication can occur where none of the machines, some of the machines, or all of the machines include the VUI of the present invention.

**[0130]** Using the standard VUI, a voice controlled device can locate other voice controlled devices within a communications environment in a number of ways. These include overhearing a human interact with another machine, overhearing a machine interact with another machine, explicitly requesting nearby machines to identify themselves by using the identification phrase "(silence) What is out there?", explicitly seeking a specific class of machines (e.g. all clocks) by addressing them by a name category "(silence) Clock are you out there?", or explicitly seeking a specific machine (e.g. a clock named Socrates) by addressing it by name "(silence) Socrates are you out there?".

**[0131]** In the first two cases, the process of listening to other conversations would reveal the other machines' names. In the other three cases the machines within earshot who respond to the "are you out there" command would respond with their names. In the last two cases, the "What is out there?" command is restricted to certain classes of machines and a specific named machine thereby limiting the number of machines that will respond to the command. Once the name of the target voice controlled device is known, the initiating voice controlled device can issue other commands (e.g. "Socrates what time is it?") to the other.

**[0132]** In some cases, a voice controlled device may need to talk to another voice controlled device, one or both of which may not adhere to the above protocol. In these cases, the machines can be explicitly programmed to issue the correct commands and recognize appropriate responses. A simple example of this interaction would be a voice controlled device with voice recognition capability and a telephone voice interface dialing a voice-based service such as a spoken report of the time, and simply capturing the desired data (the time).

**[0133]** The preferred embodiments of the present invention for METHOD AND APPARATUS FOR STANDARD VOICE USER INTERFACE AND VOICE CONTROLLED DEVICES are thus described. While the preferred embodiments of the present invention utilize a speaker independent voice recognition system, the present invention is also compatable with speaker dependent voice recognition systems. While the present invention has been described in particular embodiments, the present invention should not be construed as limited by such embodiments, but rather construed according to the claims that follow below.

the communicated appliance name and the command are communicated using audible speech.

**10.** The method of claim 8 for activating a voice controlled device, wherein,

the communicated appliance name and the command are communicated using non-audible speech.

**11.** A method of controlling a voice controlled device, the method comprising:

providing a voice controlled device having a speech recognition system for recognizing speech;
storing a default appliance name into the voice controlled device;
communicating a communicated name and a command to the voice controlled device; and
controlling the voice controlled device if the communicated name is recognized as matching the default appliance name and the command is recognized by the voice controlled device.

**12.** The method of claim 11 for activating a voice controlled device, wherein,

the communicated appliance name and the command are communicated using audible speech.

**13.** The method of claim 11 for activating a voice controlled device, wherein,

the communicated appliance name and the command are communicated using non-audible speech.

**14.** A method for activating a voice controlled device, the method comprising:

providing a voice controlled device having a speech recognition system for recognizing speech;
storing a default appliance name into the voice controlled device;
storing at least one user assignable appliance name into the voice controlled device;
communicating a communicated name and a command to the voice controlled device; and
controlling the voice controlled device if the communicated name is recognized as matching the at least one user assignable appliance name or the default appliance name and the command is recognized by the voice controlled device.

**15.** A method of assigning a new name to a voice controlled device, the method comprising:

providing a voice controlled device having a speech recognition system for recognizing speech;
activating the voice controlled device; and
communicating a new name to the voice controlled device at least once.

**16.** The method of claim 15 for assigning a new name to a voice controlled device, wherein,

the voice controlled device is activated by communicating a current appliance name and a change name command.

**17.** The method of claim 15 for assigning a new name to a voice controlled device, wherein,

the new name is communicated using audible speech.

**18.** The method of claim 15 for assigning a new name to a voice controlled device, wherein,

the new name is communicated using non-audible speech.

**19.** The method of claim 15 for assigning a new name to a voice controlled device, wherein:

the voice controlled device includes prompting capability and the voice controlled device communicates audible prompts to a user in order to request communication from the user of the new name.

**20.** The method of claim 15 for assigning a new name to a voice controlled device, wherein:

a security means to protect each voice controlled device from unauthorized use.

29. The first voice controlled device of claim 24 capable of operating in a communication environment with at least one other voice controlled device, the first voice controlled device further comprising:

a security means to protect each voice controlled device from unauthorized use.

30. The first voice controlled device of claim 27 capable of operating in a communication environment with at least one other voice controlled device, the first voice controlled device further comprising:

a security means to protect each voice controlled device from unauthorized use.

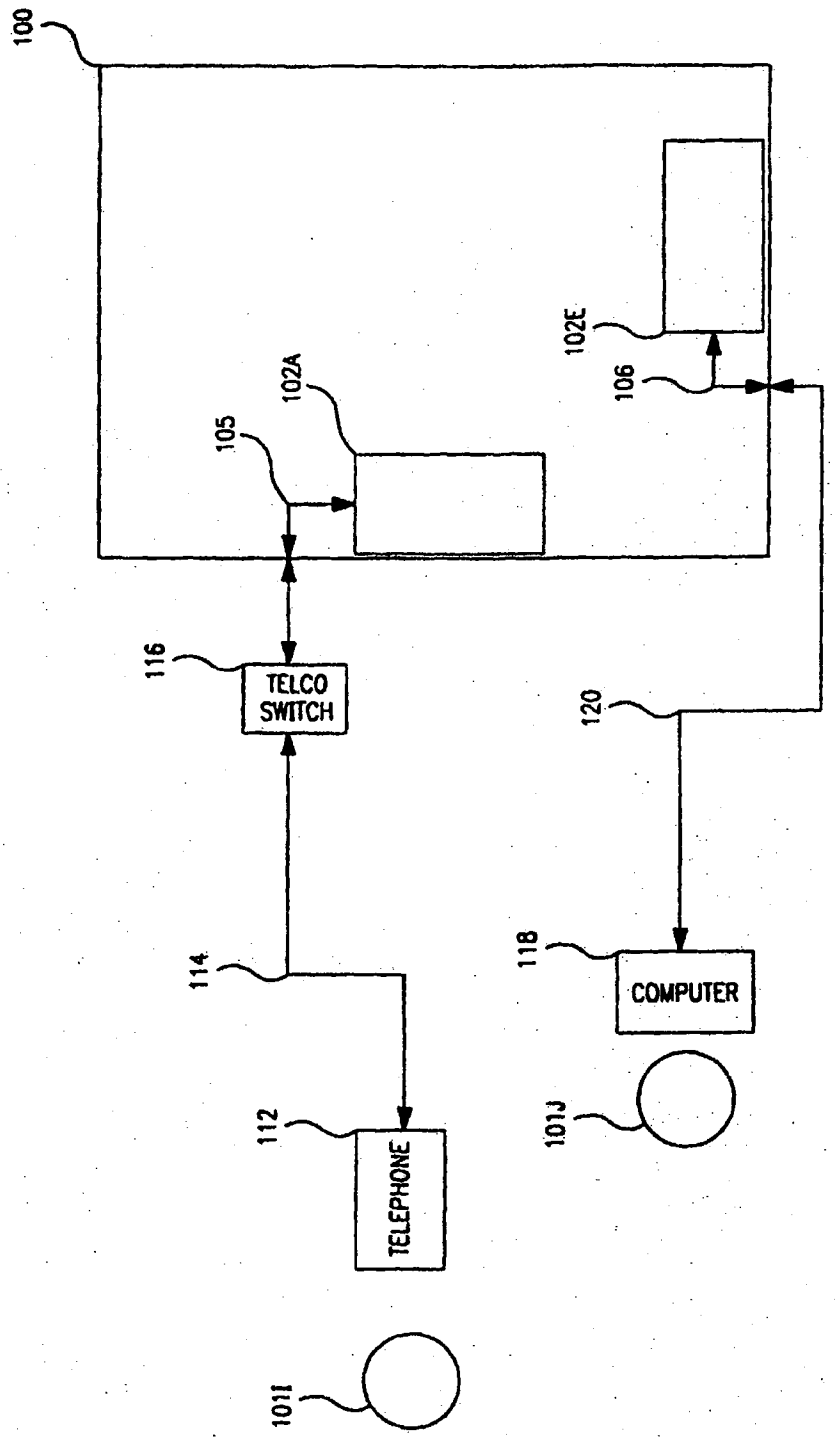
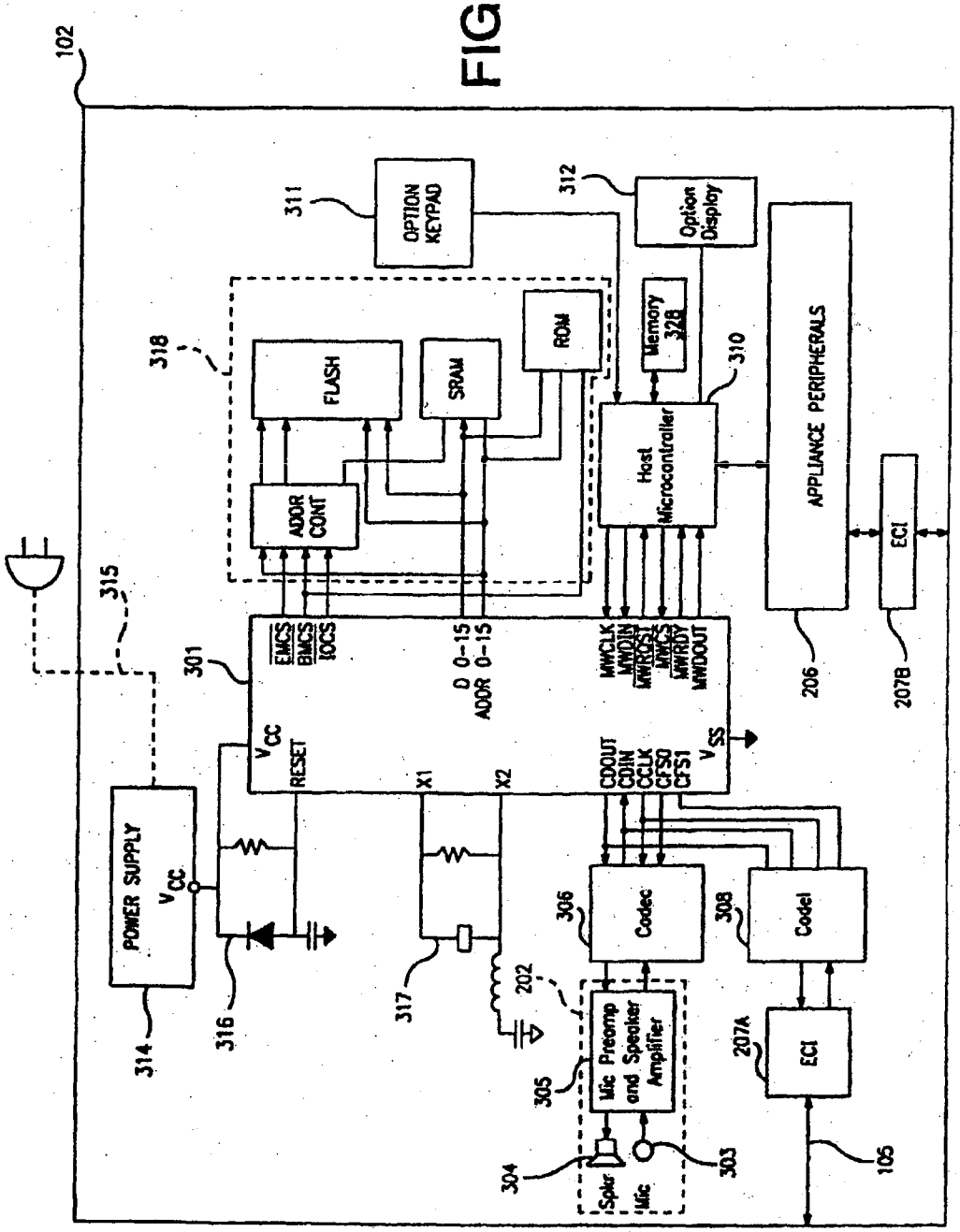



FIG. 1B

FIG. 3



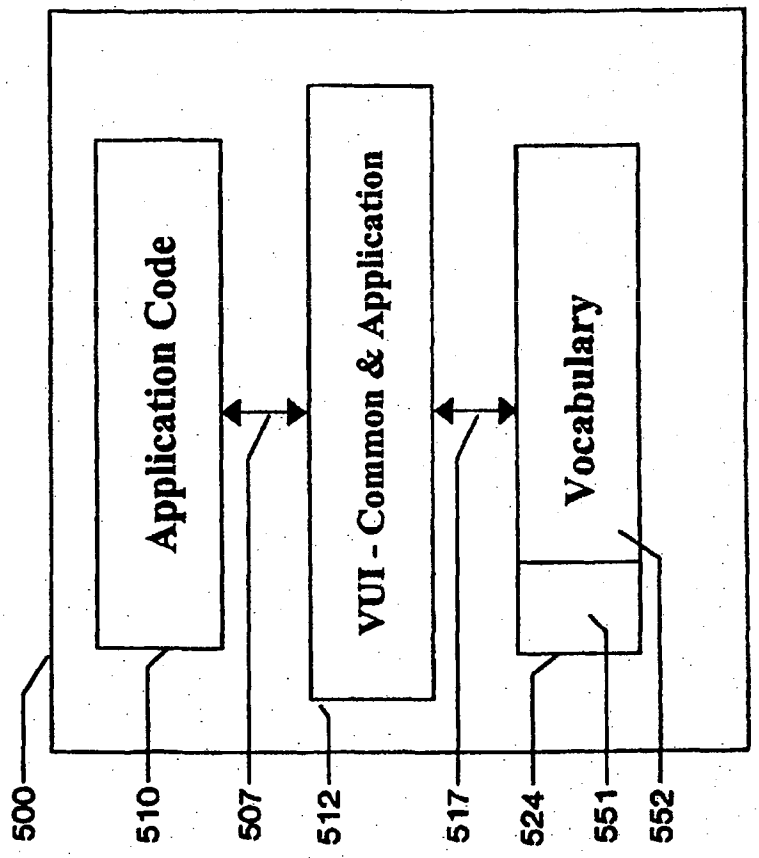



FIG. 5

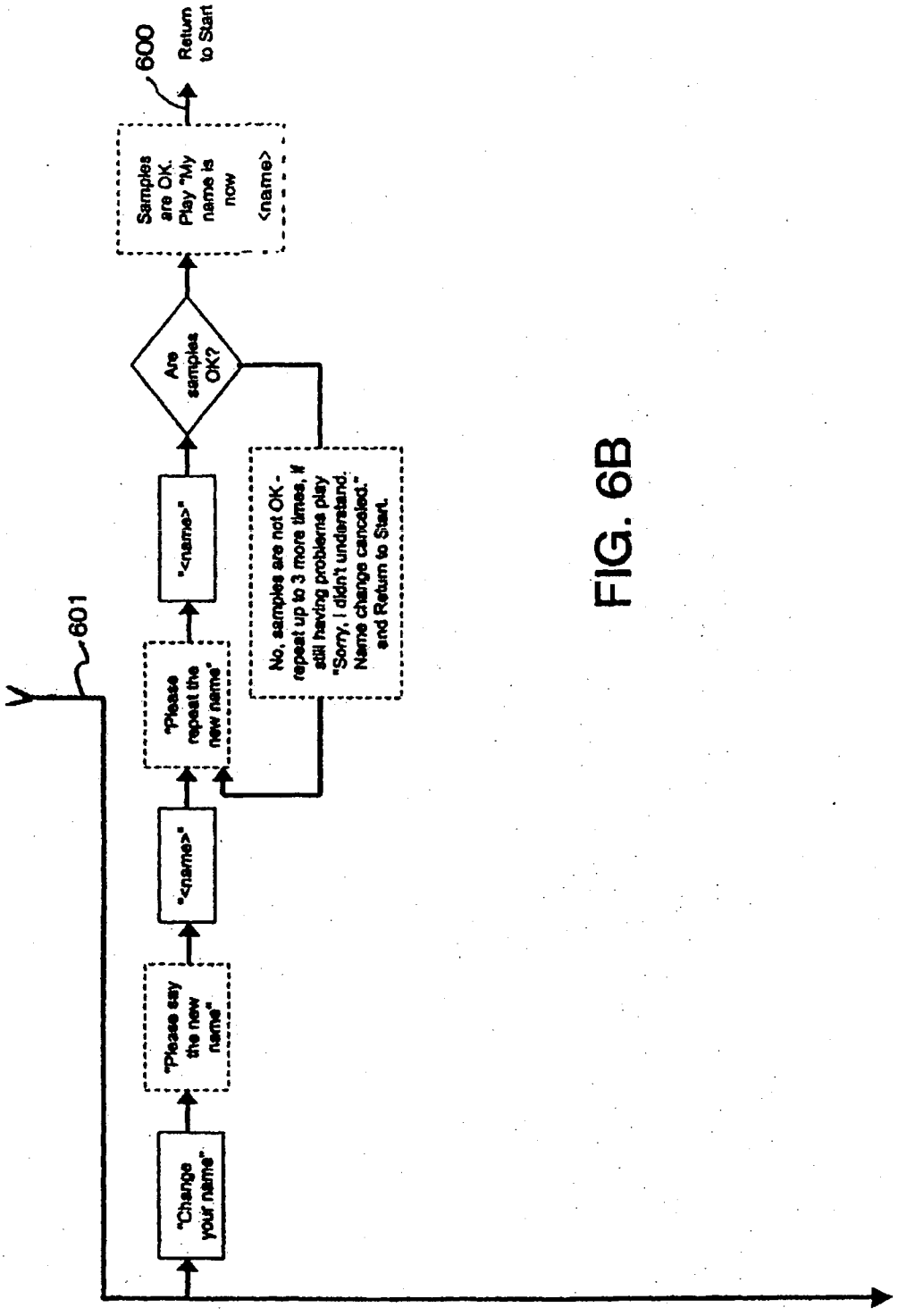



FIG. 6B

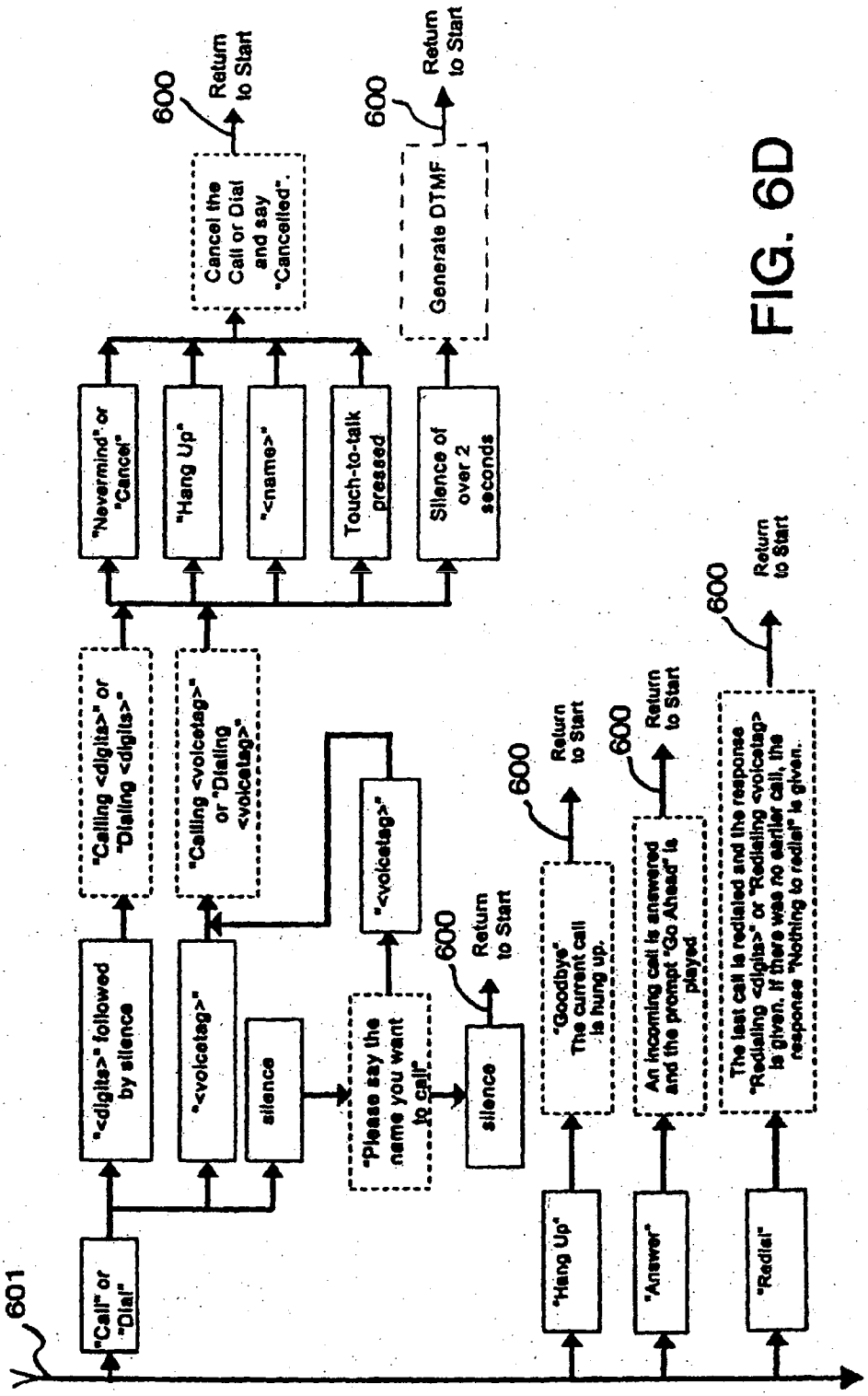



FIG. 6D

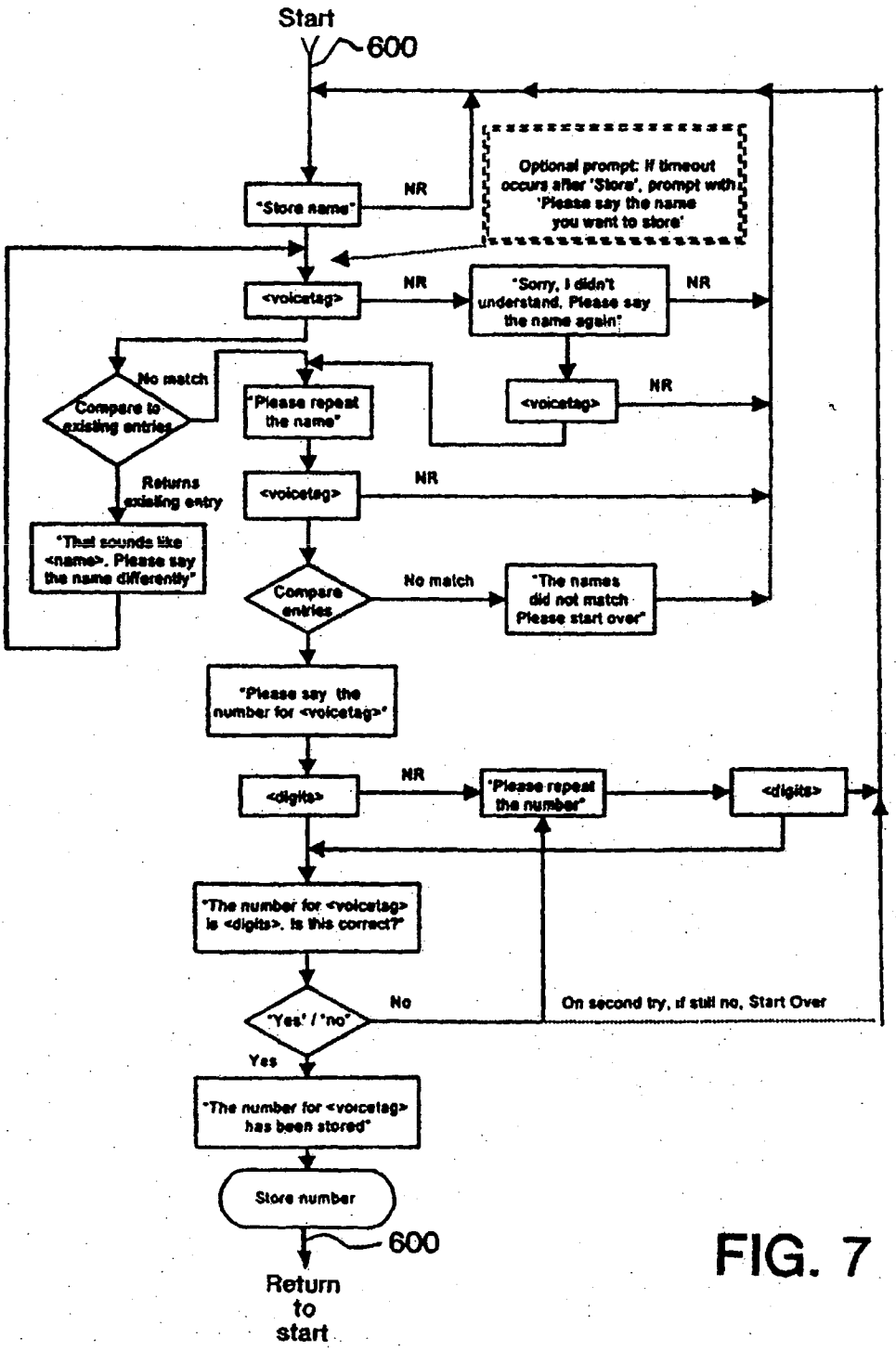



FIG. 7

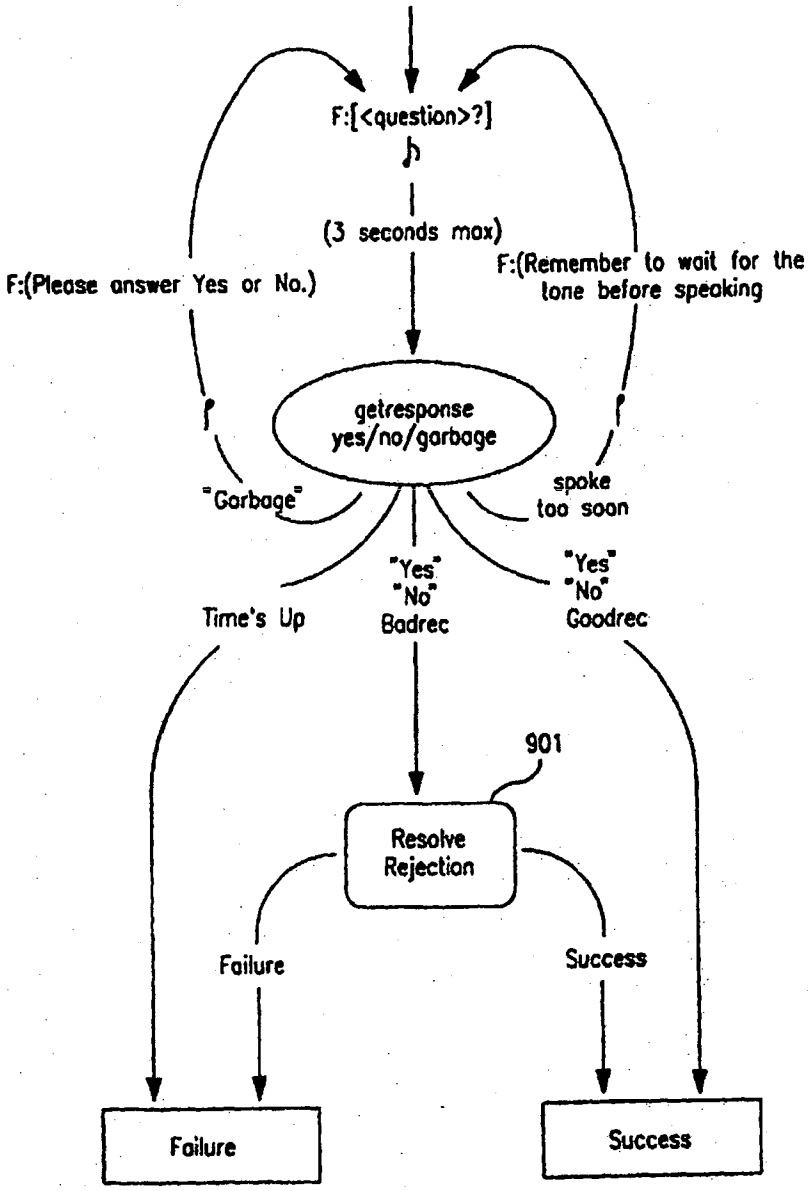



FIG. 9A

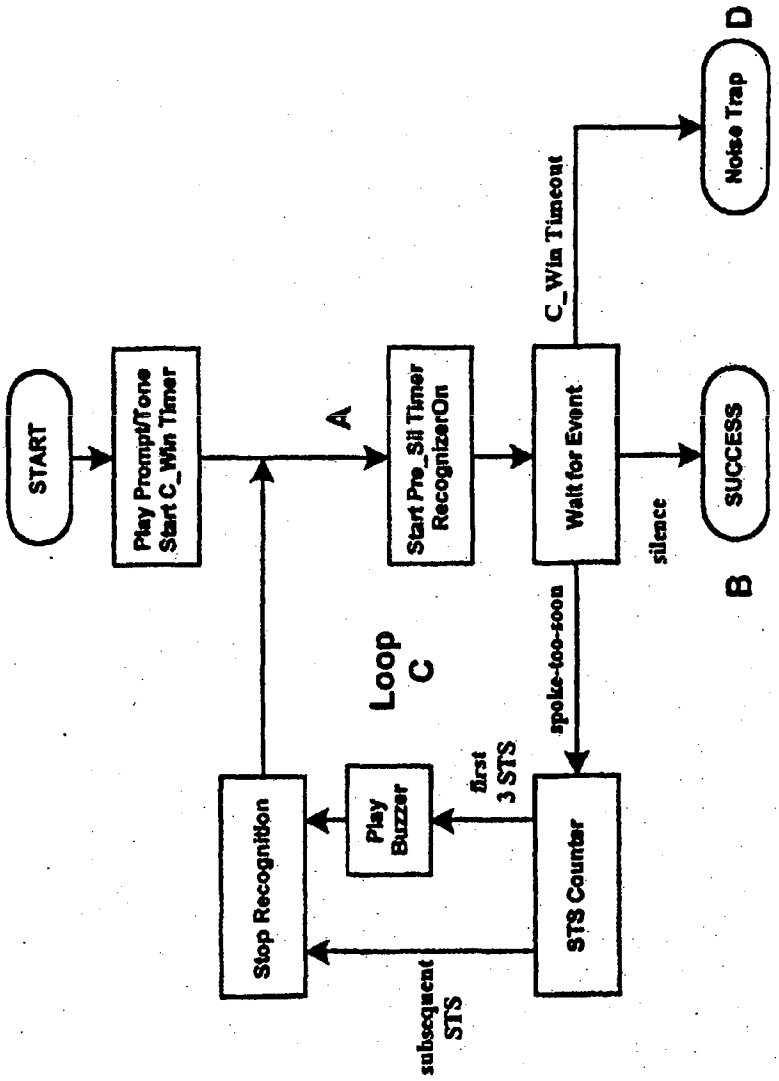



FIG. 10A

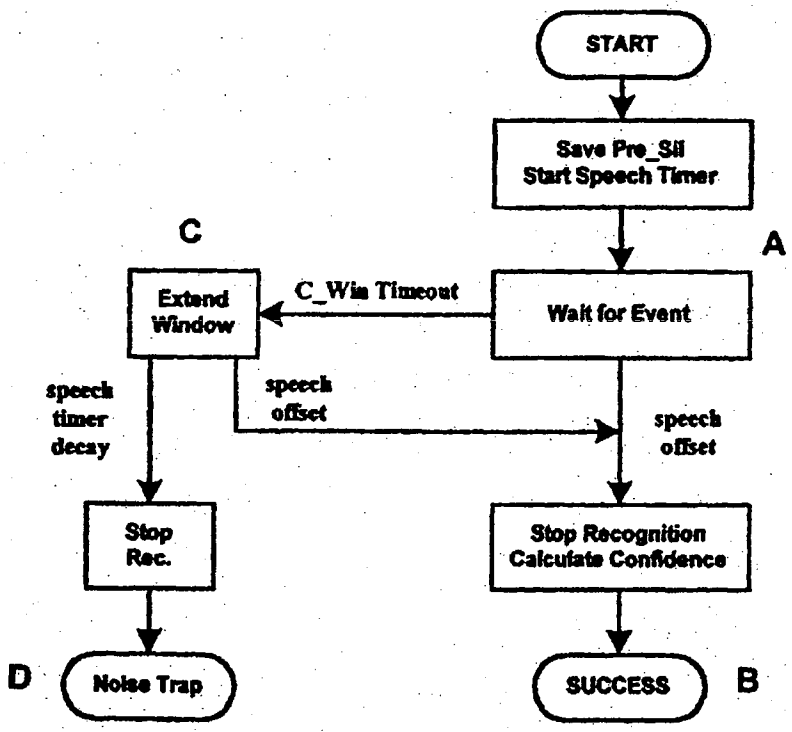



FIG. 10C

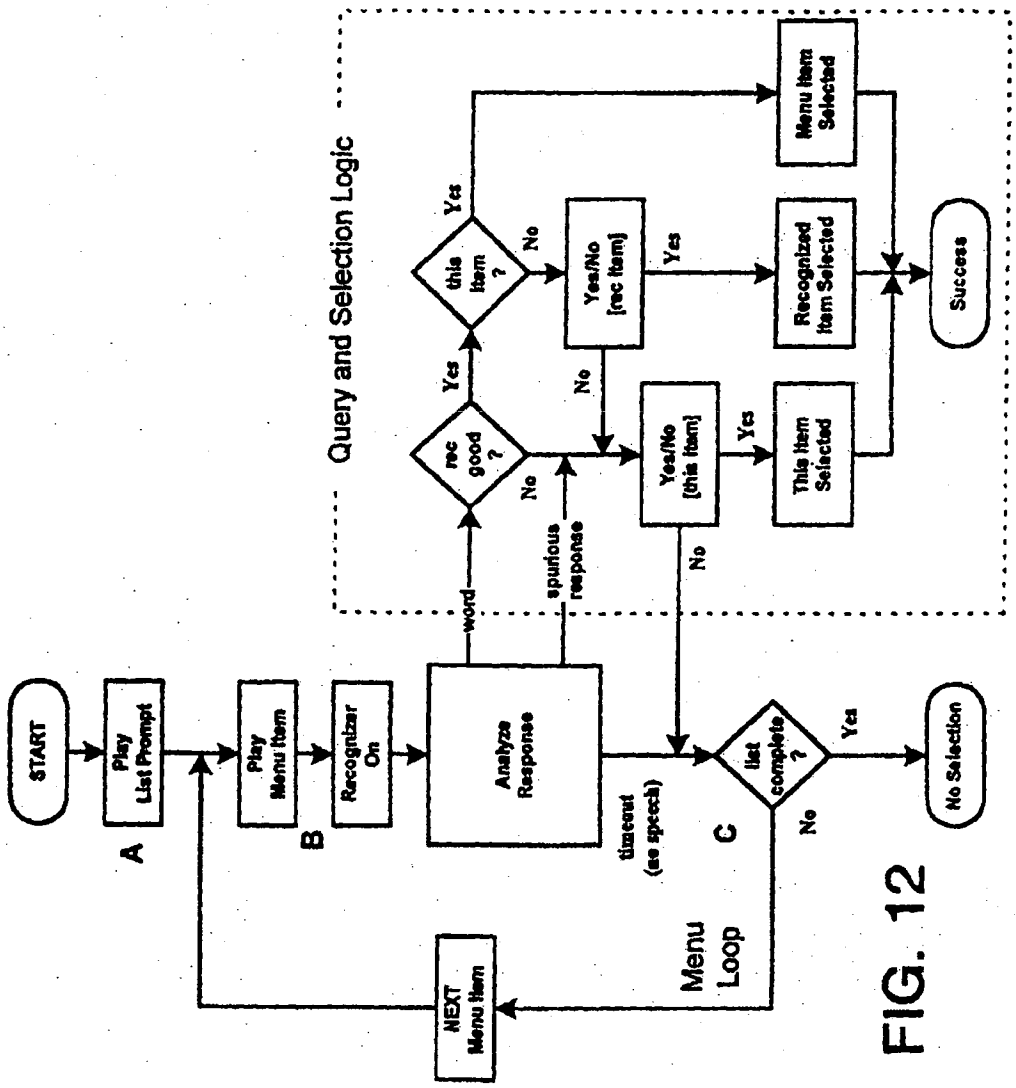



FIG. 12